Remarks on the monotonicity of default probabilities

Dirk Tasche*†

July 23, 2002

Abstract

The consultative papers for the Basel II Accord require rating systems to provide a ranking of obligors in the sense that the rating categories indicate the creditworthiness in terms of default probabilities. As a consequence, the default probabilities ought to present a monotonous function of the ordered rating categories. This requirement appears quite intuitive. In this paper, however, we show that the intuition can be founded on mathematical facts. We prove that, in the closely related context of a continuous score function, monotonicity of the conditional default probabilities is equivalent to optimality of the corresponding decision rules in the test-theoretic sense. As a consequence, the optimality can be checked by inspection of the ordinal dominance graph (also called Receiver Operating Characteristic curve) of the score function: it obtains if and only if the curve is concave. We conclude the paper by exploring the connection between the area under the ordinal dominance graph and the so-called Information Value which is used by some vendors of scoring systems.

KEYWORDS: Conditional default probability, score function, most powerful test, Information Value, Accuracy Ratio.

1 Introduction

In its new attempt – the so-called Basel II Accord – to provide quantitative rules for the capital banks are charged with for their credit risks, the Basel Committee requires banks to determine *default probabilities* for all obligors in their credit portfolios. These default probabilities can be derived from internal rating systems (Basel Committee, 2001). As a consequence, there is a growing need to develop internal rating systems in order to meet the requirements by the Basel II Accord.

However, the Basel II Accord does not only allow internal ratings but also gives rules for properties of the rating systems which will be checked by and then by the supervisory authorities. Hence, the rating systems used by the banks have to meet certain quality standards. These

*Deutsche Bundesbank, Postfach 10 06 02, 60006 Frankfurt am Main, Germany
E-mail: dirk.tasche@bundesbank.de

†The opinions expressed in this article are those of the author and do not necessarily reflect views shared by Deutsche Bundesbank.

standards include the actual state of the system as well as the process of its development. As a consequence, a lot of work has been done by several researchers in order to establish a common standard based on reasonable economic and statistical assumptions.

Monotonicity of the rating system in the sense that better ratings should correspond to lower default probabilities is one of the most important requirements (see e.g. Fritz and Popken, 2002; Krahnen and Weber, 2001). In case that the rating system is based on a score function Fritz and Popken (2002) even demand monotonicity at score level. In the paper at hand, we show that this requirement can be based on a decision-theoretic foundation. We prove that monotonicity of the conditional default probabilities is equivalent to optimality of the corresponding decision rules in the test-theoretic sense.

This paper is organized as follows. In Section 2 we present the mathematical framework and some basic statistical facts on score functions. Section 3 gives in Proposition 3.3 the main result. In Section 4 we discuss the connection between two important summary statistics for the discriminatory power of score functions: the Information Value (IV) and the Accuracy Ratio (AR).

2 Assumptions and basic facts

In the sequel we describe the result of the scoring process by a real random variable (or statistic) S on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. A second statistic T (for type) with values in the set $\{D, N\}$ (D for default and N for non-default) indicates whether the firm which has been scored will be insolvent or solvent by a previously fixed time horizon. As, however, the value of T can be observed only with some delay the financial institution faces the problem to infer its value from the known value of S.

In order to describe formally the problem we fix some assumptions and notations:

- Write short D for the event $\{T = D\}$ and N for $\{T = N\}$.
- Denote the *overall default probability* by p, i.e. $p = \mathrm{P}[D] \in (0, 1)$.
- Assume that S has a conditional density $f(s|t)$ given T, i.e.

  $$\mathrm{P}[S \in A \,|\, T = t] \;=\; \int_A f(s|t)\, ds$$

  for both $t = D$ and $t = N$ and any Borel subset A of the real line. For the sake of brevity we write

  $$f_D(s) = f(s|D) \qquad \text{and} \qquad f_N(s) = f(s|N).$$

- Given the conditional densities f_D and f_N of S for the two possible values of T, we denote the conditional distribution functions of S given the values of T by F_D and F_N respectively,

i.e.

$$F_D(s) = \mathrm{P}[S \leq s \,|\, D] \;=\; \int_{-\infty}^{s} f_D(x)\,dx,$$
$$F_N(s) = \mathrm{P}[S \leq s \,|\, N] \;=\; \int_{-\infty}^{s} f_N(x)\,dx. \tag{2.1}$$

Assumption 2.1 (Smoothness of model) *The densities f_D and f_N are positive continuous functions in some open interval $I \subset \mathbb{R}$. For any $x > 0$ the set*

$$U_x \;=\; \{s \in I : \, x\, f_N(s) = f_D(s)\}$$

has Lebesgue measure 0.

Note that under Assumption 2.1 both F_D and F_N are continuously differentiable and strictly increasing functions. In particular, their inverse functions F_D^{-1} and F_N^{-1} respectively exist and are uniquely defined.

Definition 2.2 (Likelihood ratio) *We call*

$$L(s) \;=\; \frac{f_D(s)}{f_N(s)}, \quad s \in I, \tag{2.2}$$

the likelihood ratio *at score s.*

From Assumption 2.1 follows that the conditional distributions $\mathrm{P}[L \circ S \in \cdot \,|\, D]$ and $\mathrm{P}[L \circ S \in \cdot \,|\, N]$ of the likelihood ratio applied to the score statistic given the type of the firm under consideration are continuous. To see this note that e.g.

$$\mathrm{P}[L \circ S = x \,|\, D] \;=\; \mathrm{P}[S \in U_x \,|\, D] \;=\; \int_{U_x} f_D(s)\,ds \;=\; 0$$

for arbitrary $x > 0$.

A further conclusion from Assumption 2.1 is a well-known formula (e.g Kullback, 1959, p. 4) for the *conditional default probability* $\mathrm{P}[D \,|\, S = s]$ of a firm given that its score equals s, namely

$$\mathrm{P}[D \,|\, S = s] \;=\; \frac{p\, f_D(s)}{p\, f_D(s) + (1-p)\, f_N(s)} \;=\; 1 - \big(1 + \tfrac{p}{1-p}\, L(s)\big)^{-1}. \tag{2.3}$$

Since for all $s \in I$ we have $\mathrm{P}[S = s] = 0$, this conditional probability has to be understood in the non-elementary sense (cf. Durrett, 1996, ch. 4).

If we assume that insolvent firms in general receive lower scores than solvent firms, a very simple procedure for a test of the hypothesis $T = N$ against the alternative $T = D$ is to fix a score level $s_0 \in I$ and to reject $T = N$ whenever $S < s_0$. Of course, in case of higher scores for the insolvent firms the hypothesis $T = N$ should be rejected whenever $S > s_0$. In the sequel, we will restrict ourselves to the consideration of the first case since for most of the results the transfer to the second case is obvious.

The level s_0 should be chosen in such a way that a maximal level of the *Type I error* (i.e. to reject $T = N$ although it is true) is guaranteed. For a given level $u \in (0, 1)$ this can be achieved by setting

$$s_0 = F_N^{-1}(u). \tag{2.4}$$

The classic way to compare the discriminatory power of different decision statistics is to fix the Type I error at some level and to look at the size of the *Type II error* (i.e. $T = N$ is not rejected although it is false). In the context of the test procedure specified by (2.4) the size of the Type II error is $\mathrm{P}[S \geq s_0 \,|\, D] = 1 - F_D(s_0)$. For convenience, in the sequel we will call tests of this kind *cut-off* tests. Of course, it does not matter whether the Type II error is minimized or $F_D(s_0)$ (often called the *power* of the test) is maximized over all possible statistics S. By convention, we consider here the maximization variant.

The function $\delta : (0, 1) \to [0, 1]$ defined by

$$\delta(u) \;=\; \delta_S(u) \;=\; F_D\big(F_N^{-1}(u)\big), \qquad u \in (0, 1), \tag{2.5}$$

maps every possible level of the Type I error on the power of the corresponding cut-off test. If for two test statistics S_1 and S_2, we have $\delta_{S_1}(u) \geq \delta_{S_2}(u)$ for any u then we know that S_1 delivers uniformly more powerful cut-off tests than S_2. Plotting the graphs of δ_{S_1} and δ_{S_2} provides a convenient way to check this.

In the literature, the graph of δ is known as *ordinal dominance* graph or as *receiver operating characteristic* graph (see e.g. Bamber, 1975). There is a well-known connection between the function δ and the likelihood ratio defined by (2.2), namely

$$\frac{d\,\delta(u)}{d\,u} \;=\; L\big(F_N^{-1}(u)\big), \qquad u \in (0, 1). \tag{2.6}$$

As already noticed by Bamber, from (2.6) follows that δ is concave in u if and only if the likelihood ratio L is non-increasing in s (or, by (2.3), equivalently the conditional default probability is non-increasing in s). Similarly, δ is convex in u if and only if L and the conditional default probability are non-decreasing in s.

Note that tests for $T = N$ against $T = D$ need not necessarily be of cut-off form. In general, it seems reasonable to allow for all tests which are specified by some *rejection range* $R \subset \mathbb{R}$. Such a test would reject $T = N$ if $S \in R$. In case of a cut-off test, R is described by

$$R \;=\; (-\infty, s_0). \tag{2.7}$$

As an example for other sensible forms of the rejection range, consider the case when insolvent firms tend to receive scores close to 0 whereas solvent firms achieve negative or positive scores with large absolute values. In this case the choice $R = (s_l, s_u)$ for some numbers $s_l < 0 < s_u$ appears more appropriate than (2.7). As a consequence, comparing this score function with a score function which assigns low values to insolvent and high values to solvent firms by means of the function δ from (2.5) would give a biased impression.

How can this problem be revealed? In Section 3, we will show that function δ_S is concave if and only if the S-based most powerful tests of $T = N$ against $T = D$ are cut-off. Hence, a non-concave δ_S (and equivalently non-monotonous conditional default probabilities) would indicate that the information which is contained in the statistic S is not optimally exploited.

3 Monotonous conditional default probabilities and optimal cut-off tests

For a proper formulation of the connection between conditional default probabilities and optimal cut-off tests we need a bit more of mathematical notation.

Definition 3.1 (Randomized test) *Let statistics S and T like in Section 2 with values in I and $\{N, D\}$ respectively be given. Then any measurable function $\phi : I \to [0, 1]$ is called S-based (randomized)* test *for $T = N$ against $T = D$.*
An S-based test ϕ is called test at level $\alpha \in (0, 1)$ *if* $\mathrm{E}[\phi \circ S \,|\, N] \leq \alpha$.
An S-based test ϕ^* is called most powerful test at level $\alpha \in (0, 1)$ *for $T = N$ against $T = D$ if it is a test at level α and for any S-based tests ϕ at level α we have*

$$\mathrm{E}[\phi^* \circ S \,|\, D] \;\geq\; \mathrm{E}[\phi \circ S \,|\, D]. \tag{3.1}$$

We interpret the value $\phi(s)$, $s \in I$, of a randomized test as the probability of success which should be applied in an additional Bernoulli experiment conducted by the user of the test in order to decide whether $T = N$ should be rejected. Hence, if e.g. $\phi(s) = 0.6$, the user would throw a coin with success probability 0.6 and would reject $T = H$ in case of success only. If $\phi(s) = 1$, the hypothesis has to be rejected unconditionally. In case $\phi(s) = 0$ it must not be rejected. Tests at level α are just those tests which guarantee a maximal level α of the Type I error (i.e. to reject $T = N$ although it is true). Most powerful tests at level α are those tests which minimize the Type II error size among all the tests at level α.

Randomized tests are not widespread in practice since it appears very strange to throw a coin in order to decide about the rejection of a hypothesis. However, the common deterministic tests are a subclass of the randomized tests (they are just those tests with values in $\{0, 1\}$ only), and the notion of randomized test is quite convenient for the formulation of a complete theory of test optimality.

Our main result will be based on the Neyman-Pearson Fundamental Lemma. We quote it here in a form adapted to Assumption 2.1 in order to avoid some technical difficulties. See Witting (1978, ch. 2.7) for more general versions.

Theorem 3.2 (Neyman-Pearson Fundamental Lemma)
Fix a Type I error level $\alpha \in (0, 1)$. Then, under Assumption 2.1, an S-based test ϕ for $T = N$ against $T = D$ is most powerful at level α if and only if for Lebesgue-almost $s \in I$

$$\phi(s) \;=\; \begin{cases} 1, & \text{if } L(s) > L_\alpha \\ 0, & \text{otherwise,} \end{cases} \tag{3.2}$$

where L is the likelihood ratio defined by (2.2) and L_α is any constant such that $\mathrm{P}[L \circ S > L_\alpha] = \alpha$.

From Theorem 3.2 we know that there is a deterministic most powerful test at level α, namely the test with rejection range $R = L^{-1}\big((L_\alpha, \infty)\big)$.

If the likelihood ratio L is non-increasing then the test from (3.2) has cut-off form as described in Section 2, i.e. there is a constant s_α such that the rejection range is $(-\infty, s_\alpha)$. Hence, from the observations in Section 2 it is clear that concavity of the ordinal dominance graph implies that the most powerful S-based tests for $T = N$ against $T = D$ have cut-off form. However, this concavity is not only a sufficient but even a necessary condition for the most powerful tests to be cut-off.

Proposition 3.3 (Optimality of cut-off tests)
Under Assumption 2.1, there is for every Type I error level $\alpha \in (0,1)$ *a most powerful* S*-based test with rejection range of the form* $(-\infty, s_\alpha)$ *for some* $s_\alpha \in \mathbb{R}$ *if and only if the ordinal dominance function* δ_S *as defined by (2.5) is concave. Similarly, there is for every Type I error level* $\alpha \in (0,1)$ *a most powerful* S*-based test with rejection range of the form* (s_α, ∞) *for some* $s_\alpha \in \mathbb{R}$ *if and only if the ordinal dominance function* δ_S *as defined by (2.5) is convex.*

In the Appendix, we provide a proof for the statement that existence of most powerful cut-off tests as in the first part of Proposition 3.3 implies concavity of the ordinal dominance function.

Note the dependence on the underlying statistic S in Proposition 3.3. Both the most powerful tests as well as the ordinal dominance function δ_S are defined in terms of S and its conditional distributions. As a consequence, there might be another statistic S^* such that there are S^*-based tests that are more powerful than the corresponding S-based tests at the same Type I error level. Thus, Proposition 3.3 gives a statement on the optimal use of available information when the score function has been fixed. The process of finding an appropriate score function is not subject of the proposition.

4 Information value and the area under the ordinal dominance graph

In section 3, we have investigated which conclusions can be drawn from concavity or non-concavity of the ordinal dominance graph. In this section, we will compare the area under the ordinal dominance graph as a performance measure for score functions with the so-called Information Value to be explained below.

The natural logarithm of the likelihood ratio defined by (2.2) is sometimes called *weight of evidence* (see Good, 1950). It is used by some vendors of scoring systems as a means to detect failures in exploiting the full information of a statistic. Observe that the usefulness of this method is underpinned by Proposition 3.3.

Of course, the weight of evidence is only a local measure of the information content of a statistic. Kullback (1959, p. 6) suggested the *Information Value* (or *divergence*) as the corresponding global measure. It is defined by

$$IV_S \;=\; \int_I \big(f_D(s) - f_N(s)\big) \log L(s)\, ds \;=\; \mathrm{E}[\log L \circ S \,|\, D] - \mathrm{E}[\log L \circ S \,|\, N]. \tag{4.1}$$

From the first representation in (4.1) follows that IV_S is always non-negative and symmetric in f_N and f_D. In order to arrive at a test-theoretic interpretation of the Information Value, observe that it can be equivalently written as

$$IV_S \;=\; \int_I \big(\mathrm{P}[\log L \circ S \le x \,|\, D] - \mathrm{P}[\log L \circ S \le x \,|\, N]\big)\, dx. \tag{4.2}$$

Hence IV_S is just the sum of the signed areas between the graphs of the distribution functions of $\log L \circ S$ conditional on $T = N$ and $T = D$ respectively. Define

$$S^* \;=\; \log L \circ S, \quad \text{and} \quad F_N^*(s) \;=\; \mathrm{P}[S^* \le s \,|\, N],\; F_D^*(s) \;=\; \mathrm{P}[S^* \le s \,|\, D],\; s \in I, \tag{4.3}$$

and assume that F_N^* is differentiable and strictly increasing such that $(F_N^*)^{-1}$ exists (and hence is differentiable, too). With the substitution $u = F_N^*(s)$, then (4.2) can be written as

$$IV_S \;=\; \int_0^1 \big(F_D^*\big((F_N^*)^{-1}(u)\big) - u\big)\frac{d(F_N^*)^{-1}}{du}(u)\, du. \tag{4.4}$$

Replacing the derivative $\frac{d(F_N^*)^{-1}}{du}(u)$ from (4.4) by the constant 1 yields the integral

$$\int_0^1 \big(F_D^*\big((F_N^*)^{-1}(u)\big) - u\big) du \;=\; \int_0^1 \big(\delta_{S^*}(u) - u\big) du, \tag{4.5}$$

with δ_{S^*} defined analogously to δ_S in (2.5). Here, the right-hand side of (4.5) measures the area between the ordinal dominance graph of the statistic S^* and the diagonal. Twice this area is a well-known performance measure for scoring systems – the so-called *Accuracy Ratio*. The area plus $1/2$ is just the average power of the S^*-based cut-off tests where the average is computed with equal weight 1 for all Type I error levels in $(0, 1)$. Recall, however, that in case of (4.5) the score function S has been replaced by the score function S^*.

By (4.4) and (4.5) we have seen that the concepts of information value and Accuracy Ratio are quite similar from the computational point of view. Nevertheless, they differ essentially in two aspects. First, the information value is calculated for the transformed statistic S^*. And second, for IV_S the difference between the ordinal dominance graph and the diagonal is weighted with the derivative of the inverse conditional distribution function of S^* given N.

Another way to express the similarity between the two concepts is to write the Accuracy Ratio $AR_S = 2\int_0^1 \big(\delta_S(u) - u\big)\, du$ as

$$AR_S \;=\; 2\,\big(\mathrm{E}[F_D \circ S \,|\, N] - \mathrm{E}[F_D \circ S \,|\, D]\big). \tag{4.6}$$

Hence, AR_S can be generated from IV_S essentially by substituting F_D for $\log L$ in (4.1). Both AR_S and IV_S can be interpreted as difference of the conditional expectations of a transformation of S given $T = N$ and $T = D$ respectively. However, the transformation by F_D is monotonous and yields values in a bounded range for AR_S whereas by Proposition 3.3 the transformation by $\log L$ is monotonous if and only there are most powerful cut-off tests. In general, there is no finite bound for the value of IV_S.

References

BAMBER, D. (1975) The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph. *Journal of Mathematical Psychology* **12**, 387-415.

BASEL COMMITTEE (2001) The Internal Ratings-Based Approach. *Consultative document. Supporting document to the New Basel Capital Accord.* `http://www.bis.org/publ/bcbsca05.pdf`

DURRETT, R. (1996) *Probability: theory and examples. Second Edition.* Duxbury Press: Belmont, CA.

FRITZ, S.G., POPKEN, L. (2002) Analysing Techniques for Validating Internal Credit Risk Rating Systems. *Working paper.*

GOOD, I. J. (1950) *Probability and the Weighting of Evidences.* Charles Griffin: London.

KRAHNEN, J.P., WEBER, M. (2001) Generally accepted rating principles: A primer. *Journal of Banking & Finance* **25**, 3-23.

KULLBACK, S. (1959) *Information Theory and Statistics.* Wiley: New York.

WITTING, H. (1978) *Mathematische Statistik.* B. G. Teubner: Stuttgart.

A Appendix

We sketch here the proof of the necessity part of Proposition 3.3, i.e. that concavity of the ordinal dominance function δ is necessary for the existence of a most powerful cut-off test of $T = N$ against $T = D$ at any level $\alpha \in (0, 1)$.

From (2.6) we know that, under Assumption 2.1, concavity of δ is equivalent to the likelihood ratio L being non-increasing. Hence, since L is positive and continuous by assumption, it suffices to show that for any $r > 0$ there is some $l_r \in I$ such that

$$L^{-1}\big((r,\infty)\big) \;=\; (-\infty, l_r) \cap I. \tag{A.1}$$

Choose an arbitrary $r > 0$ and let $\alpha = \mathrm{P}[L \circ S > r \,|\, D]$. By Theorem 3.2, the test "rejection of $T = N$ if $L \circ S > r$" is most powerful at level α. However, by assumption, there is an $s_\alpha \in I$ such that $\mathrm{P}[S < s_\alpha] \le \alpha$ and that the test "rejection of $T = N$ if $S < s_\alpha$" is also most powerful at level α. Again by Theorem 3.2, it follows that the functions

$$\tau_1(s) \;=\; \begin{cases} 1, & \text{if } L(s) > r \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad \tau_2(s) \;=\; \begin{cases} 1, & \text{if } s < s_\alpha \\ 0, & \text{otherwise,} \end{cases}$$

are Lebesgue-almost everywhere equal. As L is continuous by assumption, we obtain (A.1).